DOCUMENT RESUME

ED 306 277 TM 013 128

AUTHOR Saka, Thomas
TITLE Indicators of School Effectiveness: Which Are the Most Valid and What Impacts upon Them?
PUB DATE Mar 89
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Curriculum Evaluation; *Elementary Schools; Elementary Secondary Education; Grade 3; Grade 10; Instructional Effectiveness; Longitudinal Studies; Mathematics Achievement; Predictive Validity; *Predictor Variables; Public Schools; Reading Achievement; School Demography; *School Effectiveness; *Standardized Tests; State Programs; Testing Programs; *Test Results
IDENTIFIERS *Hawaii; *Indicators; Ranking; Stanford Achievement Tests

ABSTRACT

This study, focusing on 165 public elementary schools in Hawaii, examined the use of test scores as measures of curriculum effectiveness and identified context and input variables that may impact on these measures of effectiveness. Demographic characteristics and test data on approximately 12,000 third graders were used to obtain 165 school summaries for the 1987-88 school year. Data from 1984 and 1987 administrations of the Stanford Achievement Tests to 10th graders were also used in the analysis; these data were available for approximately 7,000 students. Regression models were used to allow comparisons of national and state data when the two samples were not equivalent due to socioeconomic and other demographic variables. Schools were ranked for effectiveness in reading and mathematics. Results indicate that rankings of only a few schools in Hawaii are significantly affected by the specific variables or techniques used in ranking. Future research should address the types of variables that should be included in statistical models; candidate variables include single-year indices, improvement over a period of time, and indices adjusted for demographic characteristics. Four data tables and one graph are included. (TJH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED306277

# Indicators of School Effectiveness: Which are the Most Valid and What Impacts Upon Them?

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

THOMAS SAKA

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

**Thomas Saka**
**University of Hawaii**

**Paper presented at the annual meeting of the American Educational Research Association meeting, March, 1989. San Francisco, CA.**

TM013128



BEST COPY AVAILABLE

In recent years there has been an increasing interest in evaluating the effectiveness of schools and school districts. The development of new techniques for evaluating the schools, however, has not kept pace with the increased interest. The recent furor over the proposed school report cards (Advertiser, 1988) in the state of Hawaii is an example. Two complex issues need to be considered when dealing with assessing effectiveness: defining effectiveness and developing a model in which it can be assessed.

The commonly utilized technique of comparing states, districts or individual schools based on achievement or aptitude test scores can adversely affect those schools with less-than-ideal population demographics. Achievement test score have been known to be related to socio-economic status, ethnicity, and level of parental education in addition to the actual "education" provided by the school.

The technique of comparing schools or states with similar characteristics appears to be one viable solution if grouping characteristics can be agreed upon. Part of the problem with making comparisons is related to the use of norm-referenced models in what should be a criterion-referenced situation. The norm-referenced model seems to be used because it allows for rankings and can identify the so called "good" schools.

A problem that occurs even before the selection of an evaluation model concerns the validity of the educational indicators to be used. The Center for Educational Statistics has produced a series of publications which address the issue (Stern & Hall, 1987). Test scores have been commonly used as indicators because they provide a high degree of objectivity, validity and reliability in comparison to other types of measures such as grade-point average or administrator or attitudinal ratings. One problem with the use of test scores was identified by Cannell (1988) who concludes that all 50 states reported being above average on nationally normed tests.

The need or desire to identify effective schools is not likely to disappear in the near future. Context and outcome indicators need to be identified before the specific techniques for the assessment of effectiveness can be developed. Whether consciously or unconsciously, schools will be ranked on their effectiveness in providing education to their students. Despite the fact that Hawaii is a single unified school system the diversity in population characteristics among Hawaii's public schools will make it a difficult task to provide a valid assessment of effectiveness. The techniques and indicators used in the assessment will no doubt affect some schools favorably and others adversely. The present study is an empirical attempt, utilizing existing data sources, to examine the use of test scores as measures of curriculum effectiveness and identifying context and input variables which may impact on these measures of effectiveness.

## METHOD

### Subjects

The study focused on 165 public elementary schools in the state of Hawaii. These elementary schools are located on four major island groups which comprise the seven sub-administrative districts. Demographic characteristics and test data on approximately 12,000 third grade students were used to obtain the 165 school summaries for the 1987-88 school year.

### Variables

School Effectiveness. The effectiveness of a school is often determined by the performance of its students on standardized achievement or aptitude tests. The U.S. Department of Education's "Wall Chart" (U.S. Department of Education, 1987), the Gannett ratings and the Council of Chief State School Officers (Honolulu Advertiser, 1985) use standardized test scores as major measures of success. The major problems associated with test scores for comparisons are the lack of common tests, common metric and the inconsistency of data collection.

In the state of Hawaii the Stanford Achievement Test (SAT) seventh edition is annually administered in the spring to all 10th grade public high school students as part of the statewide testing program. The SAT was chosen for use as the statewide measure because the objectives of the test are closely aligned with the state's Foundation Program Objectives upon which the local curriculum is based. School-based results are reported in terms of the percentage of students who score in the above average (7-9), average (4-6) and below average (1-3) stanine groups.

For the purposes of this study two measures of effectiveness will be used. The effectiveness of a school will first be be measured in terms of the percentage of students who score in the below average stanine (1-3) group on the reading and math subtests of the SAT. Thus in this study the effectiveness of a school increases as the percentage of students scoring in the below average stanine group decreases.

A second measure of school effectiveness is the mean stanine gain on the reading and math subtests of the SAT over a two-year period. The stanine score from the fall 1984 testing was subtracted from the stanine score from the 1987 spring testing for each student and mean gains obtained for each school. There longitudinal data were available for approximately 7,000 students. Longitudinal data provide an indication of the extent to which schools are striving for improved instruction regardless of their past levels of achievement.

Context and Input Variables. Three general groups of factors have traditionally been examined and hypothesized to affect

achievement and aptitude levels: 1) student characteristics such as socio-economic status (SES), academic skills, home environment and cultural values; 2) State characteristics including unemployment rate, per-pupil expenditure and commitment to education; and 3) school system characteristics such as teacher pay, school size, teacher experience and pupil-teacher ratio.

**National Trends**

The U.S. Department of Education's "Wall Chart" (U.S. Department of Education, 1987) lists a series of state characteristics which supposedly have an impact upon the achievement and aptitude levels of students within a state. Two recent national studies have attempted to account for the variation in mean Scholastic Aptitude Test scores across states using a number of predictor variables found on the Wall Chart. Page & Fiefs (1985) and Powell & Steelman (1984) used general characteristics of states and their educational systems to explain differences across state means. The pupil-teacher ratio and average teacher salary were used to represent characteristics of the educational system in a state. General state characteristics included percent of ethnic minority, unemployment rate, median family income, per pupil expenditure, and percent of the state budget spent on education.

The only variable found to statistically significantly correlate with the Scholastic Aptitude Test scores was the percent of ethnic minority in a state's population. The correlation of -.36 ($p < .01$) shows that the greater the percent of ethnic minority in a state, the lower the scores tended to be. This finding has been consistent over a number of years (Freedle and Kostin, 1987; Olmedo, 1981).

The pupil-teacher ratio and average teacher salary were variables that were the most highly related ($r = .18$ & $.21$) to the test scores though not statistically significant. The other state characteristic variables such as unemployment rate and median family income had correlations less than .16.

A point of interest is that the percent of seniors in each state taking the Scholastic Aptitude Test was also compared with test scores. A correlation coefficient of -.85 ($p < .01$) was obtained which shows a strong tendency for the test scores to be higher as the percent of seniors in the state taking the test is lower. This finding makes identifying factors across states which contribute to the variability in scores difficult because the relationships between the factors and test scores are confounded by nonequivalent samples of test takers across states.

Friedkin and Necochea (1988) concluded that the size of a school has an impact on academic achievement. Their use of data from the California Assessment Program found that the size of schools in medium or high SES areas was positively correlated with student performance while large schools in a low SES area had an adverse effect. The rationale given was that large schools in a high SES area would provide students with greater quantity and

quality of resources and opportunities. In a low SES setting, large school size presents problems with control and coordination.

## Hawaii Situation

In the state of Hawaii factors contributing to the variability in achievement and aptitude levels can be identified through empirical analyses of school-based data. Like the national level analyses, data on a number of factors at the local level are not available for individual students. In the state of Hawaii data on factors identified at the national level are available from school profile reports.

The problem of self selection and differing percentages of students taking the Scholastic Aptitude Test in each state on the national level does not occur when examining educational achievement in Hawaii because each student in grades 3, 6, 8 and 10 is administered the Stanford Achievement test each year.

The percent of the state budget spent on education and the per-pupil expenditure used in the national studies are held constant across all schools in the state of Hawaii. Contributions of parent-teacher associations towards school equipment such as computers may be an isolated source of differential funding. Contrary to the finding of Friedkin and Necochea (1988), an analysis of school size (SIZE) and academic achievement among Hawaii's schools found no statistically significant relationships even when including SES. Tables 1 and 2 display descriptive statistics and a correlation matrix of the measures of effectiveness, student and school system characteristics.

As shown in Table 1, SES, defined as the percent of students receiving public assistance (PUBAS) or on welfare in a school correlate .53 and .50 with the reading and math scores. A study by Saka and Chin-Chance (1989) on assessing the effectiveness of schools used median family income as the indicator of SES found correlations of -.43 and -.56 between reading and math scores and the median family income.

School system factors other than school size were examined for possible impact on educational achievement in Hawaii. Data on teacher experience (EXPL5), teacher age, and teacher absences were collected. Significant relationships between achievement and teacher experience ($r = .46$) and teacher age ($r = .45$) show that these two characteristics of the school system may affect achievement levels. Due to the high relationship between teacher age and teacher experience the age variable was dropped from further analyses. The teacher experience relationship displays a tendency for schools with large numbers of relatively new teachers to have higher percentages of students with below average achievement levels. Other teacher-related characteristics such as teacher absences had no significant relationships.

Table 1. Correlation of Effectiveness and Related Measures.

| | PRBA | PMBA | RG | MG | PRET | PUBAS | %SPED | %SLEP | %NEED | SIZE | ADA | EXPL5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMBA | .83 | | | | | | | | | | | |
| RG | .30 | .14 | | | | | | | | | | |
| MG | .14 | .01 | .49 | | | | | | | | | |
| PRET | .29 | .26 | .15 | .10 | | | | | | | | |
| PUBAS | .53 | .50 | .37 | .19 | -.03 | | | | | | | |
| SPED | .24 | .22 | -.02 | -.07 | .03 | .11 | | | | | | |
| SLEP | .25 | .20 | .24 | .06 | -.06 | .25 | .03 | | | | | |
| NEEDS | .62 | .51 | .29 | .18 | .03 | .57 | .15 | .54 | | | | |
| SIZE | -.07 | .02 | -.11 | -.15 | -.17 | -.06 | -.13 | .01 | .03 | | | |
| ADA | .13 | .14 | .05 | .03 | .01 | .18 | .04 | -.11 | .05 | -.01 | | |
| EXPL5 | .46 | .41 | .20 | .08 | .45 | .27 | .07 | -.14 | .16 | -.09 | .21 | |

Table 2. Descriptive Statistics.

| Variable | Label | Mean | SD |
|---|---|---|---|
| PRBA | % Below Avg. Reading | 23.1 | 13.1 |
| PMBA | % Below Avg. Math | 18.4 | 12.0 |
| RG | Mean Reading Gain | .1 | .3 |
| MG | Mean Math Gain | .2 | .3 |
| PRET | % Retention | .1 | .1 |
| PUBAS | % Public Assistance | 10.9 | 9.2 |
| SPED | % Special Education | 8.0 | 4.4 |
| SLEP | % SLEP | 4.0 | 4.3 |
| NEED | % Language Needs | 57.4 | 18.1 |
| SIZE | Grade Level Enrollment | 85.3 | 46.2 |
| ADA | % Avg. Daily Absences | 5.9 | 2.8 |
| EXPL% | % Tchrs < 5 yrs Exp. | 23.3 | 16.5 |

Other school system characteristics such as the percent of the students in each school classified as special education (SPED) or as limited English proficiency (SLEP) obtained statistically significant correlation between .20 and .25 with the measures of effectiveness (PRBA & PMBA).

The percent of ethnic minority which was identified as a factor in affecting aptitude levels on the national level does present as clear an analysis in the state of Hawaii because of the "melting pot" effect. Instead an immigrant - local born factor may be more of an issue. Unfortunately no systematic research to date has examined this factor nor is quantitative data readily available. However, teacher observations of this possible factor have been documented among higher ability students in science and math (McKinley High, 1988).

The impact of language skills that students possess prior to receiving formal education has been identified as having a strong relationship with the achievement levels of elementary school students (Saka, 1989). As shown in Table 1 a correlation of .62 was found between the percent of students in a school with language needs (NEEDS) as identified by the EPSS program and the percent of students scoring in the below average stanine group on the reading subtest of the Stanford Achievement Test. Language needs also correlated .52 with math achievement.

Previous literature leads one to believe that SES is strongly related to a number of factors which may impact upon aptitude and achievement. Statistically significant relationships between educational achievement, as operationalized by scores on the Stanford Achievement Test, and language needs and teacher experience were still present even after controlling for the effects of SES. The results show that language needs and teacher experience are factors which impact the achievement levels of all students independent of their financial situation.

## Empirical Analyses

Statistical adjustment has often been used to allow comparison of subjects that are not equivalent on a set of characteristics. Table 3 displays the effects of regressing the percent of below average students in each school on the three demographic variables. Both regression models obtained overall statistical significance. The percent of students receiving public assistance (PUBAS), the percent of students in a school with language needs prior to receiving formal education (NEEDS), and the percent of teachers in a school with less than 5 years of experience (EXPL5) explained significant amounts of unique variation in the regression model on the percent of students in a school scoring in the below average reading and math stanines.

Table 3. Simultaneous Regression Results on Reading and Math.

| | Sum of Squares | | | |
|---|---|---|---|---|
| | PUBAS | NEEDS | EXPL5 | $R^2$ |
| Reading | 571.6 | 4084.5 | 2940.5 | .53 |
| Math | 804.3 | 1867.6 | 1954.1 | .41 |

* All values were significant at $p < .01$

The regression analyses which contain student and school system characteristics that vary among individual school show that all schools are not equal in terms of students and resources. Residuals obtained from the regression equations can be used to give an indications of the effectiveness of a school based on the students and resources with which it is given to work. The danger is that the residuals are more normative than criterion referenced. In Figure 1 a plot of each schools' residual reading scores against a measure of SES (% on public assistance) allows the identification of the effectiveness of a school based on similar levels of SES.

A second empirical analysis involves the ranking of the effectiveness of school, in reading and mathematics, based on three indices of effectiveness. The first two columns of Table 4 (appendix) display the rankings of schools based a singles year's test scores. Columns 3 and 4 show the ranking of the schools based on mean stanine gains scores of surviving cohorts of students in each school over a two-year period. Gain scores give an indication of the improvement of a school over a period of time. An obvious discrepancy between the two indices can be seen for the school that has the highest single year's test scores but the lowest mean gains over a two year period.

Columns 5 and 6 display the ranking of the reading and math residuals adjusted for percent of students receiving public assistance in a school, percent of students with language needs prior to receiving formal education and percent of teachers in the school with less than 5 years of experience. Residual analyses, while largely normative as compared to criterion referenced, give an indication of the performance of a school relative to the characteristics of its students and resources.

## DISCUSSION

The results of the different ranking procedures to assess relative effectiveness across public schools in the state of Hawaii show that a few schools are greatly affected by the specific variables and techniques utilized. While data on school climate, curriculum offerings and student satisfaction should

ultimately be included in effectiveness measures, initial usage should rely on reliable and objective measures. Preliminary use of objective, empirical data allows the identification of relationships that occur within a particular school district.

Before techniques can be chosen, further research needs to look at the types of variables which should be included in the models, and agreement must be reached on on how effectiveness is to be defined, whether it be; 1) single year indices, 2) improvement over a period of time, or 3) indices adjusted for demographic characteristics. Another alternative is to consider all four components. The purpose of this study was not to advocate the use of any specific technique or set of variables but to demonstrate how empirical data can provide insight on the effect that the different analysis techniques could have on individual schools' measured degree of "effectiveness."

## REFERENCES


Freedle, R., & Kostin, I. (1987). Semantic and Structural Factors Affecting the Performance of Matched Black and White examinees on Analogy Items from the Scholastic Aptitude Test. Princeton, NJ: Educational Testing Service.

Friedkin N. & Necochea, J. (1088). School system size and performance: A contingency perspective. Educational Evaluation and Policy Analysis, 10, 237-249.

National rating system to benefit isles schools, superintendent believes. (1985, December). Honolulu Advertiser, p. A-8.

Olmedo, E. L. (1981). Testing linguistic minorities. American Psychologist, 36. 1078-1085.

Page, E.B. & Feifs, H. (1985). SAT scores and American States: seeking for useful meaning. Journal of Educational Measurement, 22, 305-312.

Powell, B. & Steelman, L.C. (1984). Variations in State SAT Performance: Meaningful or misleading? Harvard Educational Review, 54, 389-412.

Saka, T. (1989). Relationship between language skills prior to formal schooling and educational achievement. Presented at the Hawaii Educational Research Association Meeting, January.

Saka T., & Chin-Chance, S. (1989) A comparison of four techniques for assessing school effectiveness. Pacific Educational Research Journal, 5, 49-58.

U.S. Department of Education. (1987). State Education Statistics. Washington, DC: Office of Planning, Budget and Evaluation.

**Figure 1. Residual Reading * % Public Assistance**

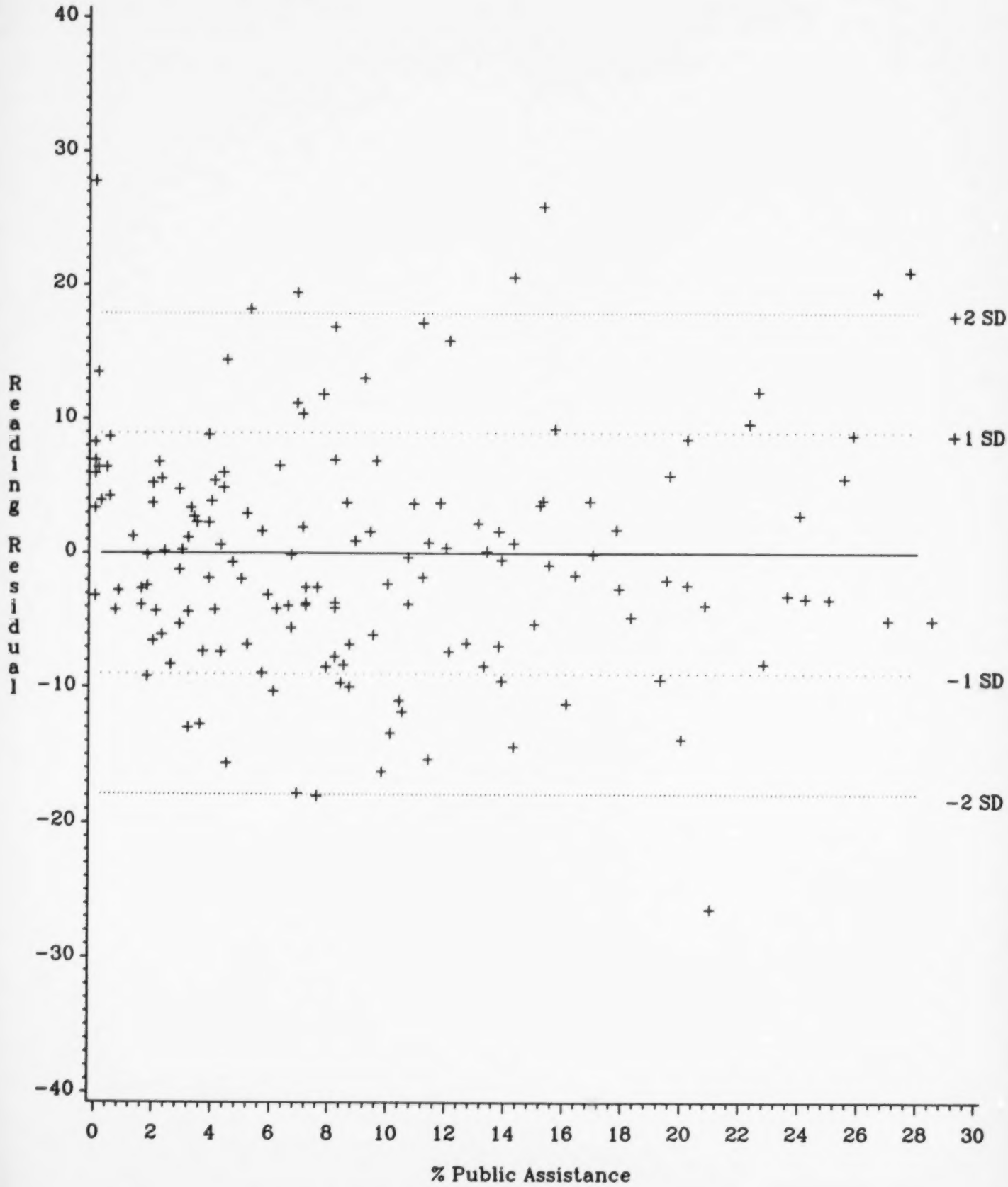

Table 4. Partial Printout of Schools Ranked by Three Methods.

| SCHOOL | (1) REGR | (2) REGM | (3) GAINR | (4) GAINM | (5) RESR | (6) RESM |
|---|---|---|---|---|---|---|
| School 1 | 1 | 57 | 165 | 164 | 10 | 103 |
| School 2 | 2 | 1 | 114 | 89 | 11 | 19 |
| School 3 | 2 | 2 | 93 | 44 | 68 | 66 |
| School 4 | 4 | 7 | 59 | 84 | 67 | 77 |
| School 5 | 4 | 23 | 78 | 118 | 49 | 92 |
| School 6 | 6 | 2 | 82 | 31 | 41 | 35 |
| School 7 | 6 | 2 | 157 | 51 | 60 | 44 |
| School 8 | 8 | 11 | 124 | 127 | 36 | 47 |
| School 9 | 9 | 15 | 141 | 32 | 32 | 45 |
| School 10 | 9 | 6 | 146 | 122 | 42 | 37 |
| School 11 | 11 | 7 | 91 | 45 | 72 | 55 |
| School 12 | 11 | 11 | 164 | 165 | 64 | 64 |
| School 13 | 11 | 2 | 43 | 58 | 3 | 5 |
| School 14 | 14 | 15 | 88 | 49 | 80 | 70 |
| School 15 | 14 | 40 | 56 | 33 | 23 | 60 |
| School 16 | 14 | 15 | 125 | 110 | 5 | 12 |
| School 17 | 14 | 15 | 6 | 25 | 8 | 11 |
| School 18 | 14 | 26 | . | . | 6 | 20 |
| School 19 | 19 | 15 | 31 | 46 | 99 | 82 |
| School 20 | 19 | 33 | 130 | 155 | 59 | 69 |
| School 21 | 19 | 15 | 62 | 147 | 28 | 26 |
| School 22 | 22 | 40 | 13 | 75 | 77 | 90 |
| School 23 | 22 | 40 | 133 | 40 | 33 | 53 |
| School 24 | 22 | 7 | 162 | 67 | 92 | 49 |
| School 25 | 22 | 23 | 118 | 156 | 50 | 52 |
| School 26 | 22 | 52 | 58 | 48 | 30 | 73 |
| School 27 | 27 | 33 | . | . | 31 | 39 |
| School 28 | 28 | 33 | 154 | 89 | 123 | 105 |
| School 29 | 28 | 61 | 129 | 81 | 40 | 94 |
| School 30 | 28 | 26 | . | . | 27 | 28 |
| School 31 | 31 | 33 | 122 | 80 | 122 | 102 |
| School 32 | 31 | 71 | 53 | 106 | 22 | 80 |
| School 33 | 31 | 47 | 3 | 25 | 7 | 22 |
| School 34 | 31 | 102 | 26 | 8 | 1 | 16 |
| School 35 | 31 | 15 | 61 | 63 | 14 | 9 |
| School 36 | 36 | 52 | 126 | 66 | 70 | 86 |
| School 37 | 36 | 33 | 106 | 152 | 110 | 84 |
| School 38 | 36 | 26 | 89 | 114 | 88 | 59 |
| School 39 | 36 | 26 | . | . | 69 | 34 |
| School 40 | 36 | 44 | 156 | 79 | 94 | 91 |
| School 41 | 41 | 91 | 104 | 105 | 115 | 145 |
| School 42 | 41 | 52 | 20 | 85 | 82 | 93 |
| School 43 | 41 | 89 | 160 | 149 | 12 | 61 |
| School 44 | 41 | 61 | 63 | 12 | 25 | 54 |
| School 45 | 45 | 57 | 138 | 41 | 51 | 76 |
| School 46 | 45 | 97 | 131 | 158 | 26 | 89 |
| School 47 | 45 | 26 | 116 | 60 | 66 | 33 |